DOCUMENT RESUME

ED 088 284 FL 004 971

AUTHOR Zelson, Sidney N. J.
TITLE Measurement and Evaluation of Speaking Skill in a Second Language.
PUB DATE [72]
NOTE 10p.

EDRS PRICE MF-$0.75 HC-$1.50
DESCRIPTORS *Achievement Tests; Audiolingual Skills; *Language Instruction; *Language Tests; Listening Tests; *Modern Languages; Phonology; *Speech Skills; Student Evaluation; Syntax; Test Construction; Test Validity; Vocabulary

ABSTRACT

A theoretical discussion of problems encountered in the measurement and evaluation of speaking skill in a second language is developed in this paper. The primary areas of interest to be evaluated are identified and discussed, including phonology, syntax, morphology, and vocabulary. A delimited and graduated method of evaluation using numerical scores is outlined in each of the target areas. (RL)

ED 088284

U.S. DEPARTMENT OF HEALTH, EDUCATION & WELFARE
NATIONAL INSTITUTE OF EDUCATION
THIS DOCUMENT HAS BEEN REPRODUCED EXACTLY AS RECEIVED FROM THE PERSON OR ORGANIZATION ORIGINATING IT. POINTS OF VIEW OR OPINIONS STATED DO NOT NECESSARILY REPRESENT OFFICIAL NATIONAL INSTITUTE OF EDUCATION POSITION OR POLICY

MEASUREMENT AND EVALUATION OF SPEAKING SKILL IN A SECOND LANGUAGE

Sidney N. J. Zelson
SUCNY at Buffalo

The testing of second language speaking skills presents a variety of problems: (1) a certain amount of equipment is often needed to administer such a test; (2) certain facilities are required if the examiner wishes to prevent one student from being influenced by another's answers; (3) whether the examinee's performance is recorded, then scored, or administered individually and scored at the same time a lack of inter-scorer reliability and intra-scorer reliability may detract from the accuracy of the measure; (4) some types of activities may go somewhat beyond testing of speaking skills only -- competency in listening and reading may be a vital factor; (5) frequently, the speaking task calls for behavior that is as dependent upon imagination and/or reasoning as it is upon linguistic or communicative competence; (6) we really do not know whether personality factors influence the examinee's performance in a speaking test to a greater degree than on other measures. If the tasks are novel, this problem may be all the more acute. There might well be two students, for example, with similar oral competence, one of whom is much more inhibited by the test situation. (7) The time element involved in scoring these measures is overwhelming.

Does a frequent formal evaluation alleviate some of these difficulties? We really don't know, but the idea seems to have

FL 004971

some face validity. Are we ready to spend the amount of time and effort necessary for such an undertaking? Is it useful to practice such tasks so that at least they will not be unfamiliar to the student? It may be a reasonable alternative that can minimize at least a few of the above-mentioned problems.

A number of decisions should be made. Do we wish to examine linguistic competence or communicative competence or both? Are they equally important? Are we interested in giving structured or unstructured tasks? How important is pronunciation? How important are other phonological features? Is phonology equal in importance to syntax, morphology, and vocabulary?

As we consider the testing of such categories as the four mentioned above, it is useful to keep in mind the unanswered questions that were previously posed. There are procedures that we might select that can help us to obtain measures that are less confounded by factors other than speaking competency. A variety of techniques may be used to test the four categories simultaneously, though phonology will be seriously slighted and diagnostically such techniques will be relatively ineffective. Yet such tasks as those to be suggested seem much more appropriate in a summative evaluation which seems to be our goal in this case.

(1) A student is told that he is to prepare to interview a visitor to our country, a visitor who speaks no English. The student is given a list of items of information he must obtain, (supplied to him in English) and is asked to record his questions.[1]

a. Where was she born?
b. When does she have to leave for her country?
c. How long has she been here?
d. Does she have any brothers or sisters?
e. What are their names?
f. Where are her parents?
g. How is she today?
h. Why doesn't she speak English?
i. Where is she going to spend the summer?
j. Where would she like to live?
k. What is her favorite season?
l. What is the weather like in the winter in her country?
m. Did the French government (Spanish, German) send her?
n. What did she do on Sundays when she was a little girl?
o. Would she stay in the United States if she could?
p. What does she want most to see in this country?
r. What did she do first upon arriving in this city?
s. Who is the tallest one in her family?
t. Does she drink wine, milk, coffee or tea with her meals?

One may observe that such questions involve a common vocabulary, a range of syntactic structures that usually differ from their English counterparts, numerous areas of pronunciation and intonation, and several tenses, though there are few items that test direct or indirect object pronouns.

If all the students were recording simultaneously, they could be given varied sequences in which to ask their questions -- perhaps each student could be instructed to start with a different number. In that way the scorer could be more confident that each tape represented the student's own work.[2]

(2) The student may be instructed to tell a story from a series of pictorial cues. He is given a few minutes for preparation.

It may be noted that the student's imagination may be a prime factor in his performance. His score may also be affected by the degree of complexity with which he expresses himself and the area

that he chooses to develop in his comments. It may happen, also, that he doesn't know vocabulary in a particular area. For those reasons it may be advisable to give him three sequences and count only the best two in his score. Provided the student has not had previous exposure to the specific context, it may be that the student's best performance gives the most valid appraisal of his speaking ability.

(3) The student may be supplied with a number of topics from which he may choose one to discuss, or he may be given several situations or roles to play, and asked to give a short monolgoue. For each of these tasks he is allowed time for preparation.

(4) The student discusses a picture at some length (with or without English cues) or makes a relevant statement about each picture of a series.

(5) The student is given one half of a dialog and asked to prepare to participate.

This type of exercise calls for reading competence, if the dialog is written in the second language. Furthermore, lack of familiarity with one lexical item in one utterance may well affect performance throughout the remainder of the test. In such a test most or all of the previously mentioned variables can come into play, either individually or in an interaction.

A common scoring procedure for evaluating such performances as those suggested involves observing four areas: (1) fluency, (2) pronunciation, (3) grammar, and (4)vocabulary. Whether or not they are weighted equally is a point for consideration, through

the literature implies that they are given equal importance. A system has appeared in recent literature, one that seems quite adaptable and that defines or delimits each graduation more clearly than several others.[3]

Pronunciation	
Phonemically accurate pronunciation throughout	4
Occasional phonemic error, but generally comprehensible	3
Many phonemic errors: very difficult to perceive meaning	2
Incomprehensible, or no response	1
Vocabulary	
Consistent use of appropriate words throughout	4
Minor lexical problems, but vocabulary generally appropriate	3
Vocabulary usually inaccurate, except for occasional correct word	2
Vocabulary inaccurate throughout, or no response	1
Structure	
No errors of morphology or syntax	4
Generally accurate structure, occasional slight error	3
Errors of basic stuucture, but some phrases rendered correctly	2
Virtually no correct structures, or no response	1
Fluency	
Speech is natural and continuous. Any pauses correspond tc those which might have been made in native language (original text reads "made by a native speaker").	4
Speech is generally natural and continuous. Occasional slight stumblings or pauses at unnatural points in the utterance.	3
Some definite stumbling, but manages to rephrase and continue.	2
Long pauses, utterances left unfinished, or no response.	1

The writer finds this scale more realistic, particularly for students at less advanced levels, than several others that have been outlined.

Speaking vocabulary may be tested separately with techniques such as response to pictorial cues, as well as with those procedures already discussed, but it may be more useful and more practicable to incorporate that category in other tests.

We may wish to evaluate the student's control of phonological features of the secord language.

a. segmental phonemes

b. suprasegmental phonemes: they may be observed separately or as <u>intonation</u>

1. stress

2. pitch

3. juncture

c. rhythm and other features

(1) The examinee listens to an utterance and repeats it. The scorer observes elements that have been predetermined but not pointed out to the examinee. Usually t , but no more than three elements may be observed per utterance.

Echo type items test the student's ability to reproduce appropriate features but not his comprehension of the principles involved, which is also a requisite for authentic speech, i.e., he may not perform similarly in a non-repetitive task.

(2) The examinee is given a series of sentences or expressions to read. Student performance is evaluated as per (1).

It is sometimes argued that orthographic symbols may cause the student to make errors where he might not in actual conversation.

(3) The student may be instructed to answer a simple question, or to respond to a simple utterance. He may be prompted with a cue -- by this means the response can be structured to a greater degree. However, this task obviously calls for comprehension also.

(4) The student may be asked to respond to pictorial cues: name the picture, tell what the person is doing, tell what time it is, what the person has in his hand, and other tasks of that type.

(5) The student may be asked to give one or more pattern drill responses. A simple substitution exercise, or another low level task, will lessen the likelihood of syntactical problems.

The scoring of the production of phonological elements will likely take one basic form, though slightly different sets of criteria may be used. If the goal is phonetic accuracy, the items might be marked on a pass/fail, native/non-native, authentic/unauthentic or with some other similar terminology; another possible treatment may be a three position scale -- phonetic accuracy-phonemic accuracy-unacceptable. The choice of one or the other would be determined by previously-established objectives. The use of acceptable/unacceptable ratings, without further definition in more precise terminology, may cause one to question the reliability of such an appraisal as well as its meaningfulness.

A student's control of morphology and syntax in a second language may be measured in the following ways.

(1) Procedures Three through Five from Phonology section (see pp. 6-7) may be adapted to the testing of morphology and syntax.

(2) Procedure One from General Speaking Test section (see pp. 2-3) may be used effectively.

(3) The student responds to pattern drills of several varieties.

It must be noted here that such an exercise may discriminate against those students who are not accustomed to using pattern drills. A more complex type of drill would aggravate the problem all the more.

(4) The student is instructed to "express the following ideas" (supplied in English). Only predetermined elements are scored.

(5) The student is supplied with a series of dehydrated sentences in the second language. Each is given with a model, so as to structure the response to the extent that the examiner wishes. The student records his responses.

The scoring of morphemic structures may be done on a correct/incorrect basis. Syntactic structures may be treated the same way or one point may be allowed for choice of the correct structure and one point for all correct forms within that structure.

Various techniques of evaluation of the speaking skills and sub-skills have been set forth and discussed. An effort has been made to outline some of the limitations, some of the variables that may detract from the validity of our measures. The primary consideration in the choice of one procedure or another has been the extent to which, in the judgment of the writer, speaking skills and only speaking skills were tested, though several tasks were included that are not pure tests of speaking ability.

While it is seldom if ever possible to eliminate all extraneous factors, it is incumbent upon the test constructor to eliminate or minimize those which he can and to be aware of those which he has not and can not. Only to that extent can we be confident of the accuracy of the information we have gathered.

FOOTNOTES

[1]Many of the questions are suggested by those in a dissertation proposal by Patricia Powell, "An Investigation of Selected Syntactical and Morphological Structures in the Conversation of High School Students After Two Years' Study of French." Additional items of this type may be found in "Directed Composition," Review Text in French Two Years (New York: Amsco Publications, 1966), p. 369.

[2]The writer has noted, in the scoring of speaking-test tapes, in installations of several kinds, that it is common to hear responses other than from student who is being scored. They often come through quite clearly, and immediately prior to the response that is to be evaluated.

[3]John L. D. Clark, Foreign Language Testing: Theory and Practice, (Philadelphia: Center for Curriculum Development, 1972), p. 93.